ORIGIN := 0

One-Way ANOVA with Random Effects

prepared by
Wm Stein

Studies involving a single classification variable with more than two levels (or states) generally fall under one-way ANOVA models. If the different levels of the classification variable have meaning in and of themselves then the variable is considered "fixed" and one may proceed with standard ANOVA questions. For instance, one may be interested in testing whether there is a general treatment effect (i.e., all mean levels equal) or whether a specific linear contrast (of means or treatment effects) is significant or not. However, if the levels of a classification variable are only a sample from a potentially large number of possibilities, then the variable is termed "random" and different questions are usually of interest. Here instead of looking at specific differences between fixed levels, one is generally more interested in estimating the amount of variance for the population from which specific levels of the study have been drawn. In more sophisticated "mixed" studies involving both fixed and random classification variables one is often interested in partitioning estimates of variance within and between levels. The example here is the simplest possible "random" (Type II) ANOVA involving only a single fixed mean level plus a single random classification variable. Data and analysis from Pinheiro & Bates (PB) 2004, *Mixed-Effects models in S and S-PLUS*.

Example:

Rail data in PB 1.1

R := READPRN("c:/DATA/Models/Rail.txt")

$travel := R^{\langle 2 \rangle}$ $\quad Y := travel$ **< response variable**

$rail := R^{\langle 1 \rangle}$ **< independent random variable**

X := READPRN("c:/DATA/Models/Rail X matrix.txt") **< cell means model contrast matrix**

$Y_{bar} := mean(Y)$ $\quad Y_{bar} = 66.5$ **< grand mean**

$$X = \begin{pmatrix} 1&0&0&0&0&0\\ 1&0&0&0&0&0\\ 1&0&0&0&0&0\\ 0&1&0&0&0&0\\ 0&1&0&0&0&0\\ 0&1&0&0&0&0\\ 0&0&1&0&0&0\\ 0&0&1&0&0&0\\ 0&0&1&0&0&0\\ 0&0&0&1&0&0\\ 0&0&0&1&0&0\\ 0&0&0&1&0&0\\ 0&0&0&0&1&0\\ 0&0&0&0&1&0\\ 0&0&0&0&1&0\\ 0&0&0&0&0&1\\ 0&0&0&0&0&1\\ 0&0&0&0&0&1 \end{pmatrix} \qquad R = \begin{pmatrix} 1&1&55\\ 2&1&53\\ 3&1&54\\ 4&2&26\\ 5&2&37\\ 6&2&32\\ 7&3&78\\ 8&3&91\\ 9&3&85\\ 10&4&92\\ 11&4&100\\ 12&4&96\\ 13&5&49\\ 14&5&51\\ 15&5&50\\ 16&6&80\\ 17&6&85\\ 18&6&83 \end{pmatrix}$$

Linear Fixed Model:

$Y_{ij} = \beta_i + \varepsilon_{ij}$ **< Cell Means model**

$Y_{ij} = \beta + \tau_i + \varepsilon_{ij}$ **< Treatments Effect model** **PB p. 6-7**

$\Sigma\tau_i = 0,\ \varepsilon_{ij} \sim N(0,\sigma^2)$

where: Y_{ij} is the response variable, β = overall mean, β_i = cell means in cell means model, τ_i = treatment effects for each group, ε_{ij} = error, with i is index of groups up to r, j = is index of replicates within each group up to n_i

Least Squares Estimation of the Regression Parameters:

$B := \left(X^T \cdot X\right)^{-1} \cdot X^T \cdot Y$

$$B = \begin{pmatrix} 54\\ 31.66667\\ 84.66667\\ 96\\ 50\\ 82.66667 \end{pmatrix}$$

^ regression coefficients

Fitted Values & Hat Matrix H:

$Y_h := X \cdot B$ **< fitted values Y_h**

$H := X \cdot \left(X^T \cdot X\right)^{-1} \cdot X^T$ **< nXn Hat matrix**

Residuals:

$e := Y - Y_h$ **< residuals**

Dimensions & Paperwork:

$n := length(Y)$ $\quad i := 0 .. n-1$ $\quad I := identity(n)$

$r := cols(X)$ $\quad r = 6$ $\quad p := r$ $\quad ii := 0 .. n-1$ $\quad J_{i,ii} := 1$

ANOVA Table:

Sum of Squares:		Degrees of Freedom:		Mean Squares:	
$SSR := Y^T \cdot \left[H - \left(\frac{1}{n}\right) \cdot J\right] \cdot Y$	$SSR = (9310.5)$	$df_R := p - 1$	$df_R = 5$	$MSR := \left(\frac{SSR}{df_R}\right)_0$	$MSR = 1862.1$
$SSE := Y^T \cdot (I - H) \cdot Y$	$SSE = (194)$	$df_E := n - p$	$df_E = 12$	$MSE := \left(\frac{SSE}{df_E}\right)_0$	$MSE = 16.167$
$SSTO := Y^T \cdot \left[I - \left(\frac{1}{n}\right) \cdot J\right] \cdot Y$	$SSTO = (9504.5)$	$df_T := n - 1$	$df_T = 17$	$MSTO := \left(\frac{SSTO}{df_T}\right)_0$	$MSTO = 559.088$

Overall F-Test for *fixed effect*:

Hypotheses:

H_0: β_i **the same** ***for all*** **i**

$\tau_i = 0$ ***for all*** **i**

H_1: ***At least one*** β_i **different**

At least one τ_i **<> 0**

< Alternate Cell means & Treatment Effects formulations of the same hypothesis

Test Statistic:

$$F := \frac{MSR}{MSE} \qquad F = 115.181$$

Critical Value of the Test:

$\alpha := 0.05$ **< Probability of Type I error must be explicitly set**

$$CV := qF[1 - \alpha, (r - 1), (n - r)] \qquad CV = 3.106$$

Decision Rule:

IF F > C, THEN REJECT H_0 OTHERWISE ACCEPT H_0

Probability Value:

$$P := 1 - pF(F, r - 1, n - r) \qquad P = 1.033 \times 10^{-9}$$

Prototype in R:

```
library(nlme)    # {nlme} for lme() & other functions
library(ape)     # {ape} for varcomp()
library(help=nlme) # prototype for finding package index
#+++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
#PINHEIRO & BATES MIXED-EFFECTS MODELS
#RAIL EXAMPLE LINEAR APPROACH
data(Rail)
plot(Rail)
Rail            #GROUPED DATA OBJECT
rail=data.frame(Rail)
rail            #CORRESPONDING DATA.FRAME

FM=lm(travel~Rail, data=rail)
summary(FM)
anova(FM)
FM2=lm(travel~Rail, data=Rail)
summary(FM2)
anova(FM2)         #lm() WORKS WITH EITHER DATA TYPE
```

Note: PB grouped data objects involve an ordering of factor levels for variable Rail. In R, this defaults to contr.poly contrasts. Conversion to unordered contrasts is therefore necessary as seen below.

```
Y=rail$travel
X=factor(rail$Rail,ordered=F) #CONVERTING TO UNORDERED FACTORS
X               #WITH TREATMENTS CONTRASTS
FM3=lm(Y~X)
summary(FM3)    #RAIL 2 IS "INTERCEPT" IN THE REPORT
model.matrix(FM3) #NOTE DIFFERENCE IN COLUMN ARRANGEMENT FOR FACTORS
anova(FM3)

FM4=lm(Y~X-1)   #CELL MEANS MODEL CONTRASTS
model.matrix(FM4)
anova(FM4)      #NOTE DIFFERENT DF FOR MODEL HERE FM4 VS FM3
```

```
> summary(FM3)    #RAIL 2 IS "INTERCEPT" IN THE REPORT
Call:
lm(formula = Y ~ X)
Residuals:
    Min      1Q  Median      3Q     Max
-6.6667 -1.0000  0.1667  1.0000  6.3333
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   31.667      2.321  13.641 1.15e-08 ***
X5            18.333      3.283   5.584 0.000119 ***
X1            22.333      3.283   6.803 1.90e-05 ***
X6            51.000      3.283  15.535 2.60e-09 ***
X3            53.000      3.283  16.144 1.67e-09 ***
X4            64.333      3.283  19.596 1.77e-10 ***
```

matches ANOVA table & F test above >

```
> anova(FM3)
Analysis of Variance Table
Response: Y
          Df Sum Sq Mean Sq F value    Pr(>F)
X          5 9310.5  1862.1  115.18 1.033e-09 ***
Residuals 12  194.0    16.2
```

df 6 here >

```
> anova(FM4)  #NOTE DIFFERENT DF FOR MODEL HERE FM4 VS FM3
Analysis of Variance Table
Response: Y
          Df Sum Sq Mean Sq F value    Pr(>F)
X          6  88911   14819   916.6 2.971e-15 ***
Residuals 12    194      16
```

```
#lmList() LINEAR MODELS FOR GROUPS
setwd("c:/DATA/Models")
K=read.table("Railframe.txt")
K
FM5=lm(travel~factor(Rail)-1,data=K)  #FM5 = CELL MEANS MODEL ANOVA
summary(FM5)
anova(FM5)  #NOTE 6 DF FOR Rail here as opposed to 5 df in FM5
lmList(travel~1|Rail,data=K) #SEPARATE REGRESSION COEFFICIENTS FOR EACH GROUP
        #GIVES CELL MEANS MODEL RESULT
railfactor=factor(K$Rail)
contrasts(railfactor)=contr.sum #CONVERTS FACTOR CONTRASTS
FM6=lm(travel~railfactor,data=K)
summary(FM6)
anova(FM6)  #df 5 for railfactor IN contr.sum MODEL ANOVA IN R
lmList(Rail) #SEPARATE REGRESSION COEFFICIENTS FOR EACH GROUP
        #GIVES CELL MEANS MODEL RESULT
```

> summary(FM5)

```
Call:
lm(formula = travel ~ factor(Rail) - 1, data = K)
Residuals:
    Min      1Q  Median      3Q     Max
-6.6667 -1.0000  0.1667  1.0000  6.3333
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
factor(Rail)1   54.000      2.321   23.26 2.37e-11 ***
factor(Rail)2   31.667      2.321   13.64 1.15e-08 ***
factor(Rail)3   84.667      2.321   36.47 1.16e-13 ***
factor(Rail)4   96.000      2.321   41.35 2.59e-14 ***
factor(Rail)5   50.000      2.321   21.54 5.86e-11 ***
factor(Rail)6   82.667      2.321   35.61 1.54e-13 ***
```

> anova(FM6) #df 5 for railfactor IN contr.sum MODEL ANOVA IN R

```
Analysis of Variance Table
Response: travel
           Df Sum Sq Mean Sq F value    Pr(>F)
railfactor  5 9310.5  1862.1  115.18 1.033e-09 ***
Residuals  12  194.0    16.2

> lmList(travel~1|Rail,data=K) #SEPARATE REGRESSION COEFFICIENTS
FOR EACH GROUP
Call:
  Model: travel ~ 1 | Rail
   Data: K
Coefficients:
  (Intercept)
1    54.00000
2    31.66667
3    84.66667
4    96.00000
5    50.00000
6    82.66667
```

Note simple way to obtain Cell Means model using lmList() >

Linear Mixed Model Approach using lme() in R:

Fitting the Least Squares regression above suffices for one-way ANOVA models using either balanced or unbalanced data. For more complex models, unbalanced data creates considerable difficulties for Least Squares model fits and estimation. A generally preferable approach utilizes Maximum Likelihood (ML) or Restricted Maximum Likelihood (REML) regressions using iterative approximation. The results (in PB's notation) are ML or REML coefficients of regression fit β for fixed effects plus "best linear unbiased predictors" (BLUP's) b for random effects. Both are generally comparable to fixed coefficients (vector B above - see final section below).

Linear Mixed Model:

$Y_{ij} = \beta + b_i + \varepsilon_{ij}$ < Random effects Model

$b_i \sim N(0,\sigma_b^2)$, $\varepsilon_{ij} \sim N(0,\sigma^2)$

In model notation:

Y ~ 1 + (1 | G)
travel ~ 1 + (1 | rail)

where:
Y = response
G = group

where: Y_{ij} is the response variable, β is overall mean (level) of the response, b_i is a random variable for levels, and ε_{ij} is random error. Levels & Errors are independent. Observations between block are independent, whereas observations within blocks are correlated.

Model Fit in R:

```
#lme() MIXED-EFFECTS MODELS APPROACH
FMe=lme(travel~1, data=Rail, random=~1|Rail)
summary(FMe)
anova(FMe)
coef(FM3)   # LEAST SQUARES FIXED COEFFICIENTS treatments MODEL IN R
lmList(Rail) # LEAST SQUARES FIXED COEFFICIENTS cell means MODEL IN R
coef(FMe)   # REML COEFFICIENTS OF FIXED EFFECTS
ranef(FMe)  # REML COEFFICIENTS OF RANDOM EFFECTS (i.e, BLUP's)
```

$$B = \begin{pmatrix} 54 \\ 31.667 \\ 84.667 \\ 96 \\ 50 \\ 82.667 \end{pmatrix}$$

```
> lmList(Rail) # LEAST SQUARES COEFFICIENTS cell means MODEL IN R
Call:
  Model: travel ~ 1 | Rail
   Data: Rail

Coefficients:
   (Intercept)
2     31.66667
5     50.00000
1     54.00000
6     82.66667
3     84.66667
4     96.00000
```

^ REML and LS coefficients differ slightly here for fixed effects

```
> summary(FMe)
Linear mixed-effects model fit by REML
 Data: Rail
      AIC      BIC    logLik
  128.177 130.6766 -61.0885
Random effects:
 Formula: ~1 | Rail
        (Intercept) Residual
StdDev:    24.80547 4.020779

Fixed effects: travel ~ 1
            Value Std.Error DF  t-value p-value
(Intercept)  66.5  10.17104 12 6.538173       0
Standardized Within-Group Residuals:
        Min          Q1         Med          Q3         Max
-1.61882658 -0.28217671  0.03569328  0.21955784  1.61437744
Number of Observations: 18
Number of Groups: 6

> coef(FMe)   # REML COEFFICIENTS OF FIXED EFFECTS
  (Intercept)
2    31.96909
5    50.14325
1    54.10852
6    82.52631
3    84.50894
4    95.74388

> ranef(FMe)  # REML COEFFICIENTS OF RANDOM EFFECTS (i.e, BLUP's)
  (Intercept)
2   -34.53091
5   -16.35675
1   -12.39148
6    16.02631
3    18.00894
4    29.24388
```

^ BLUP's represent deviations from the grand mean for each group (rail)

Variance Components from Model Fit in R:

```
#VARIANCE COMPONENTS:
VarCorr(FMe)  #SQUARE OF StdDev IN Summary.lme
varcomp(FMe)  #Alternate Function in {ape}
```

```
> VarCorr(FMe)  #SQUARE OF StdDev IN Summary.lme
Rail = pdLogChol(1)
            Variance  StdDev
(Intercept) 615.31111 24.805465
Residual     16.16667  4.020779
```

Residual errors are the same: $\sqrt{\text{MSE}} = 4.020779$

Calculation of Variance Components from ANOVA table:

$n_i := 3$ **< replicates** $MSR = 1862.1$ **< from ANOVA table above**

$a := 1$ **< fixed factor levels** $MSE = 16.167$

$$\frac{MSR - MSE}{n_i \cdot a} = 615.311$$ **< KNNL Eq. 25.50a, p. 1055**

Kutner et al. 2005 (KNNL) ***Applied Linear Statistical Models 5th Edition***

$$\sqrt{\frac{MSR - MSE}{n_i \cdot a}} = 24.805$$ **< R calculation verified above.**

Approximate Confidence Intervals in R:

```
intervals(FMe) #APPROXIMATE CONFIDENCE INTERVALS
```

confidence interval for grand mean >

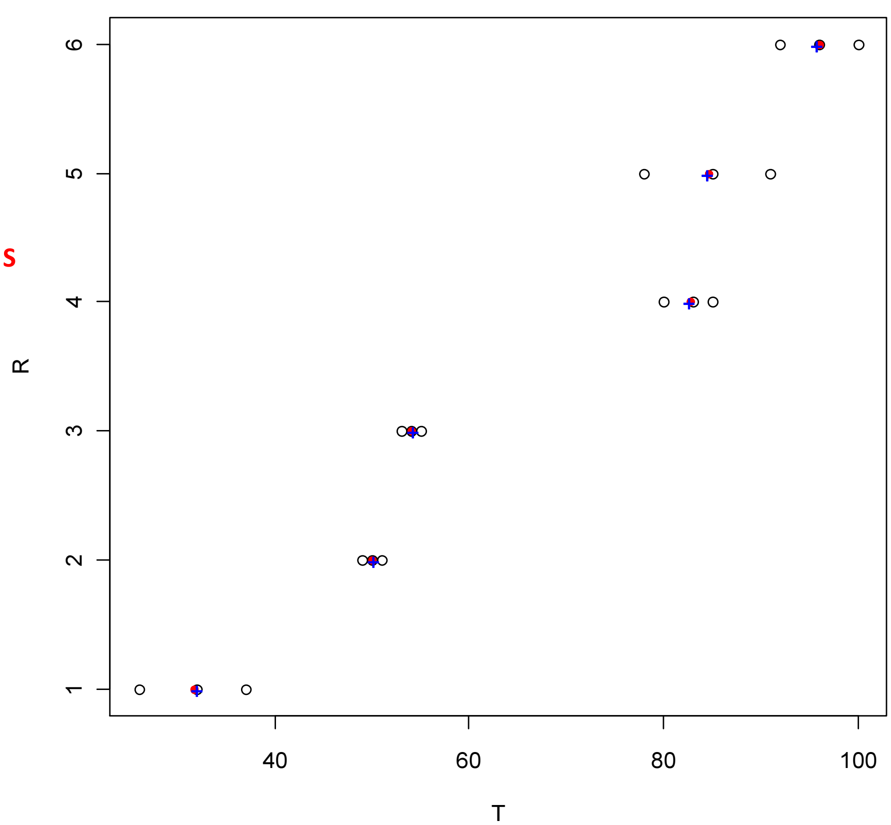
confidence interval for between group sd >

confidence interval for within group sd

```
> intervals(FMe) #APPROXIMATE CONFIDENCE INTERVALS
Approximate 95% confidence intervals
 Fixed effects:
               lower est.    upper
(Intercept) 44.33921 66.5 88.66079
attr(,"label")
[1] "Fixed effects:"
 Random Effects:
  Level: Rail
                   lower     est.    upper
sd((Intercept)) 13.27432 24.80547 46.35349
 Within-group standard error:
   lower     est.    upper
2.695000 4.020779 5.998762
```

Model Predictions and Plot in R:

```
R=Rail$Rail
T=Rail$travel
PD=predict(FM)
PE=predict(FMe)
plot(T,R,col="black")  #DATA POINTS
points(PD,R,col="red", pch=20)
#PREDICTIONS FROM lm()
points(PE,R,col="blue", pch="+")
#PREDICTIONS FROM lme()
```



Intraclass Correlation:

$\sigma := \sqrt{MSE}$ $\quad \sigma^2 = 16.167$

σ^2 is "within" (error) variance component

σ_μ^2 is "between" (level or group) variance component

$\sigma_\mu := \sqrt{\frac{MSR - MSE}{n \cdot a}}$ $\quad \sigma_\mu^2 = 102.552$

Notation follows KNNL p. 1035.

$\sigma_Y := \sqrt{\sigma_\mu^2 + \sigma^2}$ $\quad \sigma_Y^2 = 118.719$

σ_Y^2 is "total (response) variance"

$ICC := \frac{\sigma_\mu^2}{\sigma_\mu^2 + \sigma^2}$ $\quad ICC = 0.864$

< Intraclass correlation = "correlation between any two responses from the same factor level with random ANOVA model..." KNNL p. 1035

Test for overall level differences:

Note that this test is equivalent to the fixed factor test of overall effect above. Although the formal hypothesis differs, the F statistic is identical.

Hypotheses:

H_0: $\sigma_\mu^2 = 0$

H_1: $\sigma_\mu^2 <> 0$

< If $\sigma_\mu^2 = 0$ then b_i's (in PB notation above) all have to be the same

Test Statistic:

$F := \frac{MSR}{MSE}$ $\quad F = 115.181$

Critical Value of the Test:

$\alpha := 0.05$ **< Probability of Type I error must be explicitly set**

$CV := qF[1 - \alpha, (r-1), (n-r)]$ $\quad CV = 3.106$

Decision Rule:

IF F > C, THEN REJECT H_0 OTHERWISE ACCEPT H_0

Probability Value:

$P := 1 - pF(F, r-1, n-r)$ $\quad P = 1.033 \times 10^{-9}$

Least Squares Point and Interval Estimation:

KNNL pp. 1038-1047 provides least squares methods for point estimates and for confidence intervals of overall mean (β in PB = μ in KNNL), Intraclass correlation (ICC above), variance components σ^2, and σ_μ^2 (KNNL notation and above). Estimation of σ_μ^2 involves Satterthwaite or MLS procedures as explained therein. The lme() procedure in R provides similar estimates based on ML or REML fit, as explained in PB Chapter 2.

Reordering lm() results for Matrix Algebra comparison below:

$$coef_{FMe} := \begin{pmatrix} 2 & 31.96909 \\ 5 & 50.14325 \\ 1 & 54.10852 \\ 6 & 82.52631 \\ 3 & 84.50894 \\ 4 & 95.74388 \end{pmatrix} \quad \beta := \begin{pmatrix} 54.10852 \\ 31.96909 \\ 84.50894 \\ 95.74388 \\ 50.14325 \\ 82.52631 \end{pmatrix} \quad ranef_{FMe} := \begin{pmatrix} 2 & -34.53091 \\ 5 & -16.35675 \\ 1 & -12.39148 \\ 6 & 16.02631 \\ 3 & 18.00894 \\ 4 & 29.24388 \end{pmatrix} \quad b := \begin{pmatrix} -12.39148 \\ -34.53091 \\ 18.00894 \\ 29.24388 \\ -16.35675 \\ 16.02631 \end{pmatrix}$$

lme() fixed coefficients in original order ^ **lme() random BLUP's in original order ^**

Matrix Algebra Comparision of Least Squares vs lme() Models:

Least Squares:

Model: Y = XB +e

$$X = \begin{pmatrix} 1&0&0&0&0&0\\ 1&0&0&0&0&0\\ 1&0&0&0&0&0\\ 0&1&0&0&0&0\\ 0&1&0&0&0&0\\ 0&1&0&0&0&0\\ 0&0&1&0&0&0\\ 0&0&1&0&0&0\\ 0&0&1&0&0&0\\ 0&0&0&1&0&0\\ 0&0&0&1&0&0\\ 0&0&0&1&0&0\\ 0&0&0&0&1&0\\ 0&0&0&0&1&0\\ 0&0&0&0&1&0\\ 0&0&0&0&0&1\\ 0&0&0&0&0&1\\ 0&0&0&0&0&1 \end{pmatrix} \quad B = \begin{pmatrix} 54\\ 31.667\\ 84.667\\ 96\\ 50\\ 82.667 \end{pmatrix}$$

$$Y_h = \begin{pmatrix} 54\\54\\54\\31.667\\31.667\\31.667\\84.667\\84.667\\84.667\\96\\96\\96\\50\\50\\50\\82.667\\82.667\\82.667 \end{pmatrix} \quad X \cdot B = \begin{pmatrix} 54\\54\\54\\31.667\\31.667\\31.667\\84.667\\84.667\\84.667\\96\\96\\96\\50\\50\\50\\82.667\\82.667\\82.667 \end{pmatrix} \quad e = \begin{pmatrix} 1\\-1\\0\\-5.667\\5.333\\0.333\\-6.667\\6.333\\0.333\\-4\\4\\0\\-1\\1\\0\\-2.667\\2.333\\0.333 \end{pmatrix} \quad Y = \begin{pmatrix} 55\\53\\54\\26\\37\\32\\78\\91\\85\\92\\100\\96\\49\\51\\50\\80\\85\\83 \end{pmatrix}$$

Linear Mixed Model lme() in R:

Model: Y = β + Xb +ε

$Y_{bar} = 66.5$

$Yh := Y_{bar} + X \cdot b$

$\varepsilon := Y - Yh$

$$X = \begin{pmatrix} 1&0&0&0&0&0\\ 1&0&0&0&0&0\\ 1&0&0&0&0&0\\ 0&1&0&0&0&0\\ 0&1&0&0&0&0\\ 0&1&0&0&0&0\\ 0&0&1&0&0&0\\ 0&0&1&0&0&0\\ 0&0&1&0&0&0\\ 0&0&0&1&0&0\\ 0&0&0&1&0&0\\ 0&0&0&1&0&0\\ 0&0&0&0&1&0\\ 0&0&0&0&1&0\\ 0&0&0&0&1&0\\ 0&0&0&0&0&1\\ 0&0&0&0&0&1\\ 0&0&0&0&0&1 \end{pmatrix}$$

^ mean level
(β in PB notation)

$$\beta = \begin{pmatrix} 54.109\\ 31.969\\ 84.509\\ 95.744\\ 50.143\\ 82.526 \end{pmatrix}$$

^ fixed coefficients
(β_i in PB notation)

$$b = \begin{pmatrix} -12.391\\ -34.531\\ 18.009\\ 29.244\\ -16.357\\ 16.026 \end{pmatrix}$$

^ random BLUP's

$$X \cdot \beta = \begin{pmatrix} 54.109\\54.109\\54.109\\31.969\\31.969\\31.969\\84.509\\84.509\\84.509\\95.744\\95.744\\95.744\\50.143\\50.143\\50.143\\82.526\\82.526\\82.526 \end{pmatrix} \quad Yh = \begin{pmatrix} 54.109\\54.109\\54.109\\31.969\\31.969\\31.969\\84.509\\84.509\\84.509\\95.744\\95.744\\95.744\\50.143\\50.143\\50.143\\82.526\\82.526\\82.526 \end{pmatrix} \quad \varepsilon = \begin{pmatrix} 0.891\\-1.109\\-0.109\\-5.969\\5.031\\0.031\\-6.509\\6.491\\0.491\\-3.744\\4.256\\0.256\\-1.143\\0.857\\-0.143\\-2.526\\2.474\\0.474 \end{pmatrix} \quad Y = \begin{pmatrix} 55\\53\\54\\26\\37\\32\\78\\91\\85\\92\\100\\96\\49\\51\\50\\80\\85\\83 \end{pmatrix}$$